Multi Collinearity

Data set: mtcars.xlsx

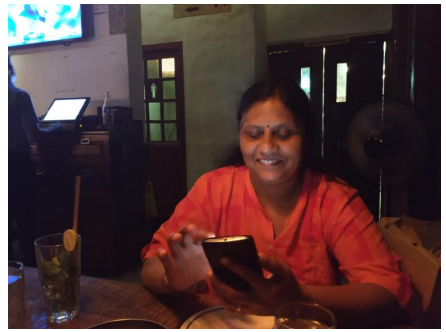
Source: Inbuilt data in R

What is Multi-Collinearity?

Dr Lylla is an amazing homemaker, a wonderful cook!

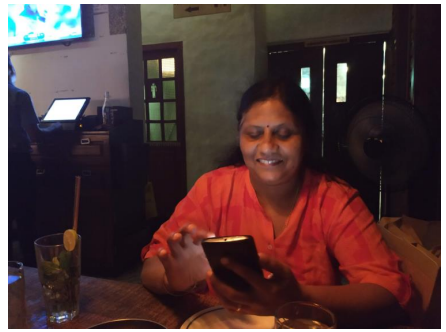
And we are inviting you for dinner, this weekend!!

What is Multi-Collinearity?
Menu – 1



Dal Makhani, Mix Dal, Dal Tadka etc.

What is Multi-Collinearity?
Menu – 2



Dal Makhani, Paneer Butter Masala, Mix Veg etc.

What is Multi-Collinearity?
Which menu you will prefer and why?

What is Multi-Collinearity?
Menu – 1










Dal Makhani, Mix Dal, Dal Tadka etc.

What is Multi-Collinearity?
Menu – 2









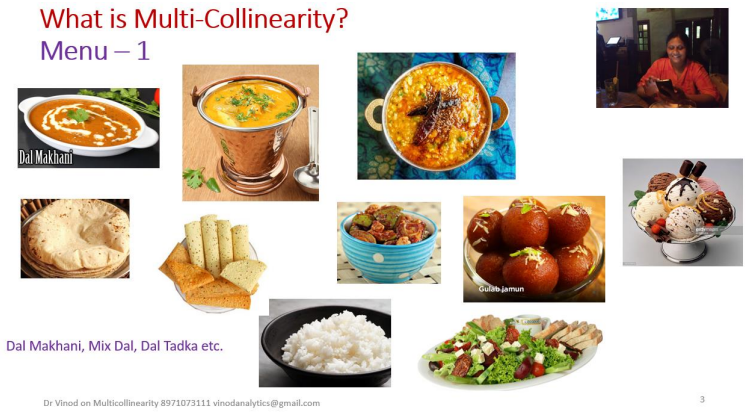


Dal Makhani, Paneer Butter Masala, Mix Veg etc.

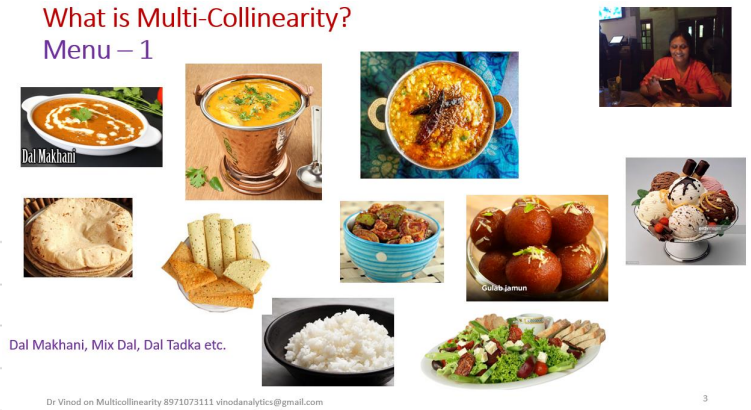
Data: Perfect Collinearity

	mpg	disp
Mazda RX4	21	160
Mazda RX4 Wag	21	160
Datsun 710	22.8	174
Hornet 4 Drive	21.4	163
Hornet Sportabout	18.7	142
Valiant	18.1	138
Duster 360	14.3	109
Merc 240D	24.4	186
Merc 230	22.8	174
Merc 280	19.2	146
Merc 280C	17.8	136
Merc 450SE	16.4	125
Merc 450SL	17.3	132

	mpg	*disp*
mpg	1	
disp	1	1



What is Multi-Collinearity? Perfect Collinearity case



	mpg	*disp*
mpg	1	
disp	1	1

SUMMARY OUTPUT

Regression Statistics	
Multiple R	1
R Square	1
Adjusted R Square	1
Standard Error	**2.79E-15**
Observations	32

=TINV(0.05,30)

2.042272456

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1126.0	1126.0	1446342305551640000000000000000000.0	0.0
Residual	30	0.0	0.0		
Total	31	1126.0			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.0	0.0	-2.0	0.1	0.0	0.0
disp	0.1	**0.0**	12026397239205200.0	0.0	**0.1**	**0.1**

Data: Near Perfect Collinearity

	mpg	disp
Mazda RX4	21	160
Mazda RX4 Wag	21	160
Datsun 710	22.8	108
Hornet 4 Drive	21.4	258
Hornet Sportabout	18.7	360
Valiant	18.1	225
Duster 360	14.3	360
Merc 240D	24.4	146.7
Merc 230	22.8	140.8
Merc 280	19.2	167.6
Merc 280C	17.8	167.6
Merc 450SE	16.4	275.8

	mpg	*disp*
mpg	1	
disp	**-0.847551379**	1

What is Multi-Collinearity? Near Perfect case

		mpg	*disp*
	mpg	1	
	disp	**-0.847551379**	1

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.847551
R Square	0.718343
Adjusted R Square	0.708955
Standard Error	**3.251454**
Observations	32

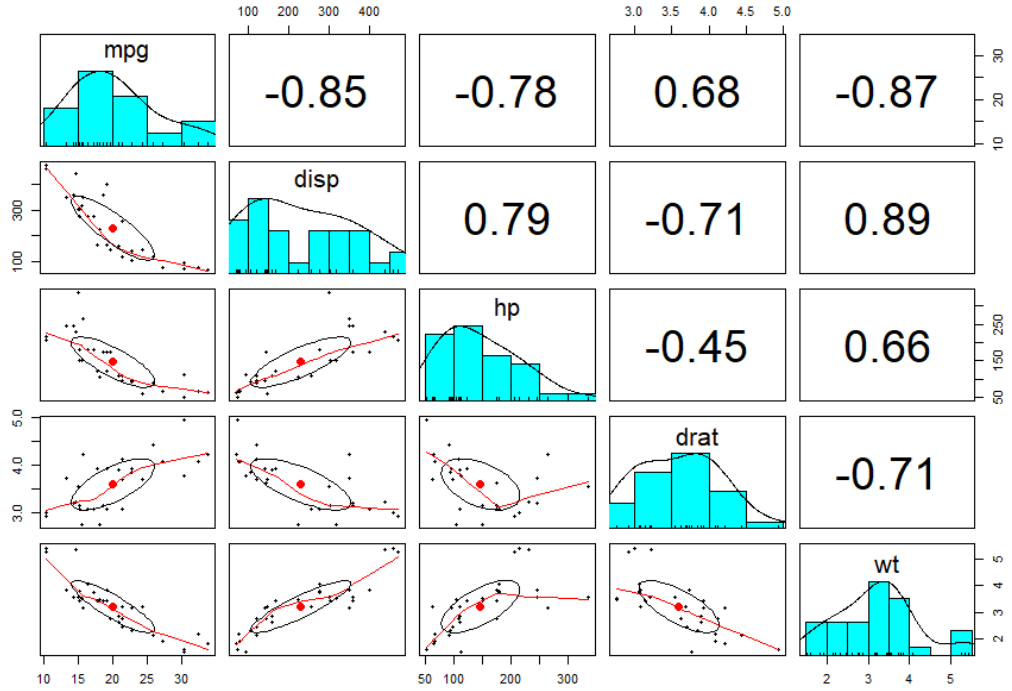
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	808.9	808.9	76.5	0.0
Residual	30	317.2	10.6		
Total	31	1126.0			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	29.59985	1.2	24.1	0.0	27.1	32.1
disp	-0.04122	**0.0**	-8.7	0.0	**-0.1**	**0.0**

How to detect Multi-Collinearity?
Data: mtcars, DV= *mpg*, IVs= *hp, disp, wt, drat*

	mpg	disp	hp	drat	wt
Mazda RX4	21	160	110	3.9	2.62
Mazda RX4 Wag	21	160	110	3.9	2.875
Datsun 710	22.8	108	93	3.85	2.32
Hornet 4 Drive	21.4	258	110	3.08	3.215
Hornet Sportabout	18.7	360	175	3.15	3.44
Valiant	18.1	225	105	2.76	3.46
Duster 360	14.3	360	245	3.21	3.57
Merc 240D	24.4	146.7	62	3.69	3.19
Merc 230	22.8	140.8	95	3.92	3.15
Merc 280	19.2	167.6	123	3.92	3.44
Merc 280C	17.8	167.6	123	3.92	3.44
Merc 450SE	16.4	275.8	180	3.07	4.07



How to detect Multi-Collinearity?
Data: mtcars, DV= *mpg*, IVs= *hp, disp, wt, drat*

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.91522066
R Square	**0.83762886**
Adjusted R Square	0.81357388
Standard Error	**2.60226088**
Observations	32

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	943.2096225	235.8024	34.82143	**2.70431E-10**
Residual	27	182.837565	6.771762		
Total	31	1126.047188			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	29.1487376	6.29358792	4.631498	8.2E-05	16.2353618	42.0621133
drat	1.76804877	1.319779454	1.339655	**0.191525**	-0.93991499	4.476012527
disp	0.00381524	0.010804852	0.353104	**0.726752**	-0.01835448	0.025984966
wt	-3.4796675	1.078371371	-3.22678	**0.003272**	-5.69230282	-1.267032241
hp	-0.0347835	0.011597337	-2.99927	**0.005756**	-0.0585793	-0.010987764

How to detect Multi-Collinearity?
Data: mtcars, DV= *mpg*, IVs= *hp, disp, wt, drat*

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.91522066
R Square	**0.83762886**
Adjusted R Square	0.81357388
Standard Error	**2.60226088**
Observations	32

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	943.2096225	235.8024	34.82143	**2.70431E-10**
Residual	27	182.837565	6.771762		
Total	31	1126.047188			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	29.1487376	6.29358792	4.631498	8.2E-05	16.2353618	42.0621133
drat	1.76804877	1.319779454	1.339655	**0.191525**	-0.93991499	4.476012527
disp	0.00381524	0.010804852	0.353104	**0.726752**	-0.01835448	0.025984966
wt	-3.4796675	1.078371371	-3.22678	**0.003272**	-5.69230282	-1.267032241
hp	-0.0347835	0.011597337	-2.99927	**0.005756**	-0.0585793	-0.010987764

High Rsqr but few significant predictors!

How to detect Multi-Collinearity?

Pair-wise correlations

Data: mtcars, DV= *mpg*, IVs= *hp, disp, wt, drat*

SUMMARY OUTPUT

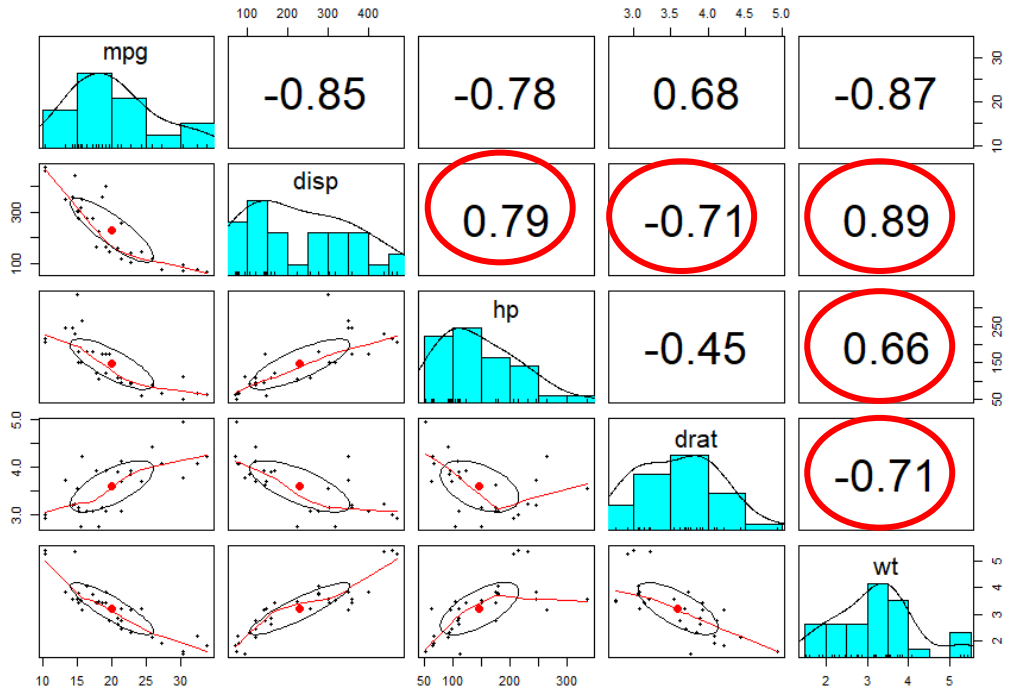
Regression Statistics	
Multiple R	0.91522066
R Square	**0.83762886**
Adjusted R Square	0.81357388
Standard Error	**2.60226088**
Observations	32

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	943.2096225	235.8024	34.82143	**2.70431E-10**
Residual	27	182.837565	6.771762		
Total	31	1126.047188			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	29.1487376	6.29358792	4.631498	8.2E-05	16.2353618	42.0621133
drat	1.76804877	1.319779454	1.339655	**0.191525**	-0.93991499	4.476012527
disp	0.00381524	0.010804852	0.353104	**0.726752**	-0.01835448	0.025984966
wt	-3.4796675	1.078371371	-3.22678	**0.003272**	-5.69230282	-1.267032241
hp	-0.0347835	0.011597337	-2.99927	**0.005756**	-0.0585793	-0.010987764

High pair-wise correlation



How to detect Multi-Collinearity?

Auxiliary Regressions

Data: mtcars, DV= *mpg*, IVs= *hp, disp, wt, drat*

	Rsqr	Num	Denom	F
disp~wt+hp+drat	0.88	0.44	0.004138	**106.3333**
wt~hp+disp+drat	0.8	0.4	0.006897	**58**
hp~disp+wt+drat	0.65	0.325	0.012069	**26.92857**
drat~hp+wt+disp	0.56	0.28	0.015172	**18.45455**
mpg~all	0.8376			
			F_Crit	**3.32765**
			=FINV(0.05,2,29)	

$$F_{x_i} = \frac{R^2_{x_i} / (k-2)}{(1-R^2_{x_i}) / (n-k+1)}$$

If F Calculated is > F Critical, multicollinearity is indicated!

OMG, all seems to be troublesome!

How to detect Multi-Collinearity?

Variance Inflation Factor

Data: mtcars, DV= *mpg*, IVs= *hp, disp, wt, drat*

```
#install.packages("HH")
library(HH)

vif(mod)
```

```
> vif(mod)
    disp       hp     drat       wt
8.209402 2.894373 2.279547 5.096601
```

Model	DV	Rsquare	VIF
1	mpg~disp+hp+wt+drat	0.837629	
2	disp~hp+drat+wt	0.878188	**8.209372**
3	hp~disp+drat+wt	0.654502	**2.894373**
4	drat~disp+hp+wt	0.561316	**2.279545**
5	wt~disp+hp+drat	0.803791	**5.096606**

$$VIF = \frac{1}{1 - R_{1.2}{}^2}$$

How to detect Multi-Collinearity?
Klien's rule of thumb
Data: mtcars, DV= *mpg*, IVs= *hp, disp, wt, drat*

Model	DV	Rsquare	VIF
1	mpg~disp+hp+wt+drat	0.837629	
2	**disp~hp+drat+wt**	**0.878188**	**8.209372**
3	hp~disp+drat+wt	0.654502	**2.894373**
4	drat~disp+hp+wt	0.561316	**2.279545**
5	wt~disp+hp+drat	0.803791	**5.096606**

If Rsqr for any predictor is > overall Rsqr, the predictor is having multicollinearity!


58+5
Never Stop Learning
abc
325
8+5
designed by freepik


Fall down seven times
Stand up eight
~Japanese Proverb